# Cyber Risk in Banking: Measuring and Predicting Vulnerability[*]

Steven D. Baker[†] Dimuthu Ratnadiwakara[‡]

Sept 16, 2025

## Abstract

We construct a bank–quarter panel linking external cybersecurity ratings, realized cyber incidents, and regulatory financial data for U.S. banks (2015–2024). Using quarterly rolling-window random forest models with dynamic feature selection, we predict whether a bank will experience a cyber incident within the next year. Both bank characteristics and cybersecurity posture provide independent and complementary predictive content, with the combined model achieving the highest out-of-sample accuracy. Predictive performance remains robust across bank sizes and forecast horizons and is not driven by simple persistence in incident history. Interpretation of the fitted models highlights the consistent importance of size, balance sheet composition, and specific security controls, along with interactions between these controls, in predicting cyber incidents.

**Keywords:** cyber risk, banking, predictive modeling

[*] The results and views are those of the authors and do not reflect those of the Federal Reserve Bank of Richmond, or the Federal Reserve System.

[†] The Federal Reserve Bank of Richmond (steven.baker@rich.frb.org)

[‡] The Federal Reserve Bank of Richmond (dimuthu.ratnadiwakara@rich.frb.org)

# 1. Introduction

Cyber risk has rapidly become one of the most pressing threats in the banking sector, commanding heightened attention from industry and regulators (Kashyap and Wetherilt, 2019; Duffie and Younger, 2019; The International Monetary Fund, 2024). Banks' heavy reliance on digital infrastructure and online services has vastly expanded their attack surface, exposing them to sophisticated cyber threats. A successful cyberattack can impair payment systems, disrupt credit provision, and undermine confidence in the broader financial system, creating channels for rapid contagion across institutions (Eisenbach, Kovner, and Lee, 2022, 2025; Kopp, Kaffenberger, and Wilson, 2017). Despite growing supervisory attention, there are no standardized measures of bank-level cyber risk that supervisors or industry can apply across the sector. Existing approaches—largely disclosure-based or market-implied (Florackis, Louca, Michaely, and Weber, 2023a; Jamilov, Rey, and Tahoun, 2023; Jiang, Khanna, Yang, and Zhou, 2024)—are designed for large publicly traded firms and cannot be readily extended to include small and mid-sized banks.

We assemble a bank-quarter panel that links externally observed cybersecurity posture, realized cyber incidents, and detailed balance sheet information representing the entire U.S. banking sector. The dataset combines BitSight technical risk assessments, Zywave incident reports, and FFIEC Call Report filings, allowing for consistent coverage of banks of all sizes from 2015 through 2024.

BitSight provides standardized indicators of network hygiene, configuration practices, and evidence of compromise, while Zywave classifies and dates cyber incidents using proprietary monitoring. Call Reports supply quarterly financial and structural characteristics for U.S. banks. Integrating these sources produces a unified framework for forward-looking analysis of cyber vulnerability, enabling sector-wide assessment of how security posture relates to future incidents and supporting the development of supervisory metrics applicable beyond large, publicly traded institutions.

We begin by presenting descriptive statistics that summarize the cybersecurity landscape of U.S. banks. These include the cross-sectional distribution of technical security indicators, the incidence and composition of cyber events over time, and systematic variation in these patterns by bank size and balance sheet characteristics. The key takeaways from this exercise are twofold: first, while most banks score highly on core security metrics, there is meaningful dispersion—particularly in certain risk vectors—that leaves subsets of institutions more exposed; and second, realized incidents are disproportionately concentrated among the largest banks, consistent with their greater digital complexity and visibility to attackers and with the key role of size in attacker incentives, as modeled in equilibrium by Ramírez (2025).

We then evaluate the out-of-sample predictive content of cybersecurity posture and bank characteristics for future incidents, where the dependent variable is a binary indicator equal to one if the bank experiences at least one incident within the next year. Using a quarterly random forest framework with dynamic feature selection—where the most predictive cybersecurity indicators are re-estimated each quarter—we find that both information sets contribute independent forecasting power. Models based solely on BitSight signals achieve an average out-of-sample AUC of 82.5, while bank-characteristics-only models reach 86.2. Combining the two yields the highest average AUC of 89.8, an improvement of over three percentage points relative to the best single-source specification. This gain underscores the complementarity between externally observed security posture and financial characteristics in predicting cyber events.

To ensure that the observed predictability is not simply capturing persistence in which institutions report or experience incidents, we augment the combined model with lagged incident history—defined as an indicator for whether the bank experienced a qualifying event in the previous four quarters. A model using only this variable performs poorly, indicating that recent history alone provides limited predictive value. When lagged incidents are added to the combined specification, the AUC rises only marginally, suggesting that most of the forecasting power comes from contemporaneous security posture and bank characteristics rather than mechanical recurrence of past events.

Next, we assess whether the model's predictive power is concentrated in specific parts of the banking sector or tied to a particular forecast window. Splitting the sample by bank size shows consistently high accuracy across small, mid-sized, and large institutions, underscoring that the combination of technical security signals and balance sheet characteristics generalizes beyond the largest, most visible banks. Likewise, performance remains strong for prediction horizons ranging from one quarter to one year, indicating that the same predictors support both near-term monitoring and longer-term supervisory assessments.

Finally, we compare our model performance to textual metrics developed in Florackis et al. (2023a) and Jamilov et al. (2023) for public banks. Our model substantially outperforms the textual metrics, achieving out-of-sample AUC averaging around 90% versus around 60% for each of the textual metrics. Adding the textual metrics to our model results in a small performance improvement of around 1% AUC on average.

Beyond demonstrating high predictive accuracy, our analysis provides new insights into the factors underlying cyber risk in the banking sector. Variable importance results reveal that bank size and balance sheet composition—particularly total assets, deposits-to-assets, and loans-to-assets—are among the most consistent predictors of incidents, rivaling or exceeding the importance of individual security indicators. On the cybersecurity side, specific features such as patching cadence, TLS/SSL configuration, TLS/SSL certificates, and DKIM records emerge as the most informative BitSight measures, while several commonly tracked indicators, including botnet infections and spam propagation, show limited relevance in recent years.

Interaction analysis further reveals that cyber vulnerability often reflects combinations of risk factors rather than isolated weaknesses. For bank characteristics, the interaction between loans-to-assets and deposits-to-assets stands out, potentially capturing broader customer relationships and a larger attack surface. Among security indicators, patching cadence consistently interacts with multiple controls—most notably TLS/SSL configuration and TLS/SSL certificates—suggesting that the co-occurrence of unpatched systems and insecure communication protocols materially elevates incident risk. These patterns underscore the value of combining operational and technical perspectives when assessing cyber resilience and point to areas where supervisory attention could yield the greatest risk reduction.

This paper contributes to three strands of literature: the study of cyber incidents as a source of operational and systemic risk in banking, the measurement of firm-level cyber risk exposures, and the policy discussion on supervisory tools for mitigating cyber vulnerabilities.

First, a growing body of research documents the substantial and persistent financial repercussions of cyber incidents. Between 2012 and 2017, major banks incurred roughly $200 billion in operational risk losses, with cyber events consistently among the most significant drivers The International Monetary Fund (2024). Empirical evidence shows that breaches can erode franchise value and impair operations well beyond the immediate remediation period. Kamiya, Kang, Kim, Milidonis, and Stulz (2021) find that firms suffering cyberattacks with customer data losses experience shareholder wealth declines far exceeding direct costs, as well as significantly lower sales growth for up to three years. Such shocks can propagate beyond the affected institution, causing payment system disruptions, reputational contagion, and heightened sector-wide risk perceptions. Chernobai, Ozdagli, and Wang (2021) show that greater business complexity increases the frequency of operational losses, while Berger, Curti, Mihov, and Sedunov (2022) demonstrate that large operational losses can raise systemic risk through both direct solvency effects and correlated losses across banks. Kopp et al. (2017) highlight how cyber risk can generate market failures that threaten financial stability, motivating supervisory standards and collective action to enhance resilience. The International Monetary Fund (2024) emphasizes that cyber risk has become a macrofinancial stability concern with systemic dimensions, noting that severe incidents can impair critical financial infrastructure, trigger liquidity strains, and generate broad confidence shocks. They stress that rising interconnectedness and reliance on digital technologies amplify the potential for contagion across institutions and borders, underscoring the need for coordinated supervisory responses and robust resilience frameworks. Our study extends this literature by documenting the incidence and drivers of cyber events across the entire U.S. banking sector, providing the first system-wide evidence that includes small and mid-sized banks absent from prior analyses.

Second, the analysis relates to research developing measures of firm-level cyber risk. Florackis et al. (2023a) construct a disclosure-based index from cyber-related language in 10-K filings, showing that it predicts future incidents and is priced in equity markets. Similarly, Jamilov et al. (2023) derive a market-based "cyber risk exposure" factor from earnings call transcripts, linking heightened cyber discussion to lower valuations, higher volatility, and sizable aggregate costs even without realized attacks. Jiang et al. (2024) employ machine learning to create a predictive cyber risk index that outperforms individual indicators and is associated with a positive risk premium. Ottonello and Rizzo (2024) focus on the software supply chain as a source of cyber risk, showing that new software vulnerabilities are a significant source of risk to which market participants react, but often slowly. While these studies advance measurement, they focus almost exclusively on large publicly traded firms with rich market or disclosure data.

Extending the scope to the banking sector, Murphy, Tindall, Klemme, Suek, and Dunbar (2025) combine external cybersecurity signals with financial structure information to estimate average annual cyber loss rates for U.S. banks. Complementing these approaches, Eling and Wirfs (2019) use operational risk data to distinguish between routine and extreme cyber event costs through actuarial and statistical techniques, identifying key drivers—such as human error—that disproportionately affect financial institutions. Jin, Li, Liu, and Nainar (2023) link banks' discretionary loan loss provisions — an indicator of internal control and risk management quality — to the likelihood of future cyber attacks, identifying an accounting-based early warning signal for cyber vulnerability. Heo (2024) applies the textual metric of Florackis et al. (2023a) banks, finding that cyber risk increases the probability of bank default. Gogolin, Lim, and Vallascas (2021) find that cyber incidents decrease branch deposit growth rates at small banks, which they attribute to reputational damage. Our work contributes along different dimensions, focusing on ex-ante risk indicators, especially those with specific technological underpinnings, that are applicable to nearly the entire US banking sector.

A distinct but related line of literature studies how banks react to the cyber risks of their customers in setting loan terms. Huang and Wang (2021) finds that firms with reported data breaches face higher loan spreads and increased likelihood of collateral requirements and loan covenants. Sheneman (2025) reaches similar conclusions but focuses on ex-ante risk, more in line with our approach. Whereas these papers study bank reactions to customers' cyber risks, we study the risks to banks themselves.

Third, the study contributes to the policy literature on cyber risk to the banking sector. On the one hand, our work supports a partial market-based solution. Evidence that commercially available risk-indicators are informative across the banking sector strengthens incentives to invest in cybersecurity, as such investments are at least partially observable to clients and counterparties who acquire the risk-indicators. Ahnert, Brolley, Cimon, and Riordan (2024) investigate such costly signals in a theoretical model, finding that they increase cybersecurity investment and welfare in equilibrium. However, they also show that regulation offers avenues to further improve the equilibrium. Kashyap and Wetherilt (2019) emphasize that cyber risk differs from other operational risks in its intentionality, potential for stealth, and rapid propagation, and argue for supervisory frameworks tailored to these features. Bouveret (2018) discuss the limitations of market discipline in addressing systemic cyber threats, advocating for macroprudential oversight and crossinstitutional contingency planning. Our analysis advances that agenda by producing a supervisory-relevant measure of bank-level cyber vulnerability that integrates both observable security posture and fundamental institutional characteristics. The resulting risk metric enables regulators to identify institutions most susceptible to future incidents, supporting targeted oversight and macroprudential risk assessment in an increasingly digital banking environment.

## 2. Data

Our analysis combines data from three primary sources: BitSight cybersecurity ratings, Zywave cyber incident reports, and regulatory bank-level financial filings from the FFIEC Call Reports. These datasets allow us to link observed security posture, realized cyber events, and institutional characteristics at the bank-quarter level. BitSight provides external assessments of cybersecurity conditions based on technical signals collected from internet-facing infrastructure. Zywave compiles detailed data on cyber incidents. Call Reports contain standardized regulatory disclosures on bank size, balance sheet composition, and profitability. Together, these sources form a panel suitable for modeling the determinants of cyber risk in the banking sector.

We use monthly BitSight cybersecurity ratings to quantify each bank's external security posture based on a standardized set of risk vectors. The data include a range of technical signals related to network hygiene, configuration practices, and evidence of compromise, which we aggregate to the quarterly level. Specifically, available indicators include measures such as patching cadence, TLS/SSL configuration, web application headers, email authentication protocols (e.g., DKIM records), and botnet infection rates, among others. These signals are derived from externally observable behaviors and vulnerabilities and are designed to capture an institution's exposure to cyber threats across multiple dimensions.[1]

[1] See here for more details: https://help.bitsighttech.com/hc/en-us/articles/

Although BitSight cybersecurity ratings have been available for roughly a decade, we are aware of little independent analysis assessing their predictive value out of sample and none covering the banking sector. The study most comparable to ours is by Choi and Johnson (2021), who evaluate BitSight ratings in relation to hospital cyber incidents. Although there are many differences in statistical methodology and focus, the overall findings are similar: BitSight scores, including some more granular scores, can be effective for rank-ordering firms by incident risk when used in combination with industry-specific firm characteristics. Caro Rincon and Ordóñez (2023) document the relationship between BitSight's overall security rating, incident rates, and distance to default for public companies.[2] Instead, we focus on banks, many of which are not public, and find that more granular BitSight risk-indicators and firm characteristics are necessary to rank-order firms within the industry. We also assess out of sample predictive performance using Zywave incidents, as opposed to contemporaneous incidents provided by BitSight.

Figure 1 plots the distribution of BitSight cybersecurity risk vector scores in 2019 Q4 separately for small (< \$1 billion), mid-sized (\$1–10 billion), and large (> \$10 billion) banks. Each panel corresponds to one of the 16 available security metrics, with higher values indicating better performance (i.e., lower observed risk) for that dimension. Most indicators—such as botnet infections, patching cadence, open ports, and TLS/SSL certificates—show distributions tightly clustered near the upper end of the scale, reflecting generally strong cybersecurity posture across banks. However, some metrics exhibit wider dispersion or pronounced lower tails, suggesting that certain institutions lag behind peers in specific security practices. For example, DNSSEC and Web Application Security display notable variation, particularly among mid-sized banks. Overall, the distributions indicate that while most banks maintain relatively high technical security scores, there remain measurable and heterogeneous weaknesses across several risk vectors, even within the same size category.

360007320574-A-Guide-to-Navigating-and-Prioritizing-Bitsight-Risk-Categories-Risk-Vectors

[2] Moody's, which produced the study, acquired an ownership stake in BitSight in 2021.

Zywave provides detailed records of cyber incidents compiled from public records and news sources using a proprietary tracking system. Each record includes the incident date, a standardized incident type (e.g., data malicious breach, phishing, spoofing, social engineering, network/website disruptions, IT configuration or processing errors, and fraudulent use or account access), and a text description indicating the severity of the incident.[3] For a subset of observations, the dataset also reports the number of affected records or direct losses (e.g., from legal settlements), offering additional granularity on breach magnitude. Each incident is associated with a unique Zywave firm identifier as well as the domain name of the affected institution, which enables us to merge the data with BitSight cybersecurity ratings and regulatory filings at the bank level.

To align the incident data with other sources, we transform the incident-level dataset into a domain-by-quarter panel. We first restrict the sample to incidents falling into categories data malicious breach, phishing, spoofing, social engineering, network/website disruptions, IT configuration or processing errors, and fraudulent use or account access. Next we aggregate events by domain and calendar quarter. For each domain-quarter observation, we construct three binary indicators that serve as forward-looking outcomes in our empirical analysis: whether the domain experienced at least one qualifying cyber incident in the next quarter, in the next two quarters, or within the next four quarters.

Figure 2 plots the quarterly number of cyber incidents at U.S. banks by incident type from 2015 through 2024. Malicious data breaches dominate throughout the sample period, accounting for the vast majority of reported events and exhibiting substantial quarter-to-quarter volatility. Other categories—including phishing and social engineering, IT configuration errors, fraudulent account access, and network disruptions—occur far less frequently and remain relatively stable over time. The persistent prevalence of data breaches, combined with the steady presence of other incident types, highlights the

[3] Zywave reports both an ”accident date” and a potentially later ”first notice date” reflecting disclosure for each incident. To ensure that our risk-indicators reflect vulnerabilities observable prior to the occurrence of an incident we use accident dates in our study.

ongoing exposure of the banking sector to a range of cyber threats and reinforces the importance of forward-looking measures to monitor and mitigate these risks.

Figure 3 plots the percentage of banks experiencing at least one incident in each quarter, segmented by bank size. Incident frequency is consistently highest among banks with more than $10 billion in assets, peaking at over 20% of institutions in certain quarters. Mid-sized banks ($1–10 billion) show moderate and somewhat volatile incident rates, while small banks (less than $1 billion) rarely report incidents, remaining near or below 2% throughout. This distribution suggests that cyber risk is disproportionately concentrated among the largest institutions, consistent with their greater digital exposure, complexity, and visibility to attackers.

We obtain regulatory financial data for U.S. banks from the FFIEC Call Reports, which provide standardized quarterly information on bank balance sheets, income statements, and off-balance-sheet exposures. From these filings, we construct a set of control variables commonly used in bank risk analysis: total assets, asset growth, return on equity, deposits-to-assets, and loans-to-assets. These variables capture key dimensions of bank size, growth, profitability, and funding structure, and are included in all model specifications. The Call Reports also contain the domain name of each reporting institution, which we use to merge the financial data with external cybersecurity and incident records.

Table 1 compares the characteristics of banks included in our final analysis sample to those excluded due to missing domain linkages, separately for each size category. Across all size groups, in-sample banks tend to be larger and display stronger profitability and balance sheet ratios than those not covered. For banks with over $10 billion in assets, in-sample institutions are not only larger on average but also report higher deposit-to-asset and loan-to-asset ratios. Among mid-sized banks ($1–10 billion), in-sample banks have higher ROE (11.0% vs. 9.5%) and loan intensity. Even among smaller banks (less than $1 billion), in-sample institutions are slightly larger and more profitable than those excluded.

The final dataset is structured as a bank-quarter panel that integrates three sources of information. Each observation includes (i) current-quarter bank characteristics from Call Reports, (ii) current-quarter BitSight cybersecurity signals, and (iii) forward-looking incident indicators from Zywave, defined over one-, two-, and four-quarter horizons. This unified panel allows us to examine the relationship between observed security posture and future cyber incidents while controlling for underlying institutional characteristics.

# 3. Cross-Bank Metric of Cyber Risk Exposure

This section develops and evaluates a predictive model of cyber incidents at the bank level. We adopt a random forest classifier as our primary modeling approach. This method is well-suited for capturing complex, non-parametric relationships and allows for flexible interaction structures without requiring explicit specification. Later, Section 4 characterizes the main drivers of our model. We begin by conducting feature selection to identify the most informative security signals from the BitSight dataset, then estimate the random forest model using historical incident data from Zywave. Model performance is evaluated based on the area under the receiver operating characteristic curve (AUC), with all results reported using out-of-sample predictions.

## 3.1. Feature Selection

While BitSight provides an overall "Security Rating" intended to summarize an organization's cyber risk posture, we do not use this composite score as our primary predictive variable. For U.S. banks in our sample, the Security Rating alone exhibits limited discriminatory power for future incidents, and its proprietary weighting obscures the relative importance of underlying risk vectors. From a supervisory perspective, understanding which specific security dimensions contribute to elevated risk is essential for designing targeted interventions and monitoring emerging vulnerabilities. Our approach therefore focuses on modeling the granular BitSight feature set directly, allowing us to identify the most informative predictors and assess their stability over time.

To identify the most informative cybersecurity signals for predicting cyber incidents, we implement a sequential, out-of-sample feature selection procedure using quarterly rolling windows. In each quarter q beginning in Q1 2017, we train a random forest model on all available historical data prior to q (e.g., the Q1 2017 model is trained on data from Q1 2015 through Q4 2016) and evaluate it on bank observations in quarter q. The model includes a set of core control variables—log assets, asset growth, return on equity, deposits-to-assets, and loans-to-assets—which are included in all specifications. These variables are motivated by prior variable importance analysis and serve as fundamental controls for bank size, growth, profitability, and funding structure. As documented in earlier sections, incident rates vary meaningfully with bank size and balance sheet composition, justifying their inclusion as baseline covariates.

In each training window, we extract variable importance scores from the fitted random forest and identify the top five features among those not included in the required bank controls. These selected variables represent the most predictive cybersecurity-specific signals—such as patching cadence, insecure ports, or evidence of malware—based on their contribution to out-of-sample classification accuracy. By repeating this procedure for each quarter starting in Q1 2017, we allow the feature set to vary over time as the informativeness of different signals evolves. The selected features for each quarter are then used in the final predictive model described in the following subsection.

Table 2 reports the top selected cybersecurity features for periods between 2017 Q2 and 2023 Q4. Several indicators appear consistently, suggesting their enduring relevance for banks' cyber vulnerability. DKIM is repeatedly selected, reflecting the importance of maintaining robust email authentication to limit exposure to phishing and spoofing attempts. TLS/SSL configuration and certificate management recur across periods, consistent with the critical role of secure communication channels for online banking platforms and other internet-facing services. Web Application Security also appears frequently, indicating that vulnerabilities in customer-facing or operational portals remain a persistent concern. Patching cadence is a prominent predictor in later years, underscoring the value of timely remediation of known vulnerabilities, especially in complex banking IT environments.

Other signals, such as Botnet Infections and Potentially Exploited, appear only in specific intervals, which may partly reflect overfitting in the early, smaller sample period rather than true shifts in threat activity. While some of these episodic features may capture genuine changes in the cyber landscape, their limited persistence cautions against treating them as stable indicators. This reinforces the importance of focusing supervisory and internal monitoring on consistently predictive measures, while treating transient predictors as context-dependent signals that require ongoing validation before incorporation into long-term risk frameworks.

## 3.2. Random Forest Model

To assess the predictive value of bank characteristics and externally observed cybersecurity signals, we estimate a sequence of quarterly random forest models over the 2017 Q1–2024 Q4 period. Each model predicts whether a bank will experience a qualifying cyber incident within the subsequent four quarters. We evaluate three predictor sets: (i) bank characteristics alone, (ii) BitSight cybersecurity indicators from the feature-selection procedure, and (iii) the combination of both. Out-of-sample performance is measured using the area under the receiver operating characteristic curve (AUC), with higher values indicating greater ability to distinguish between banks that will and will not experience a future incident.

Figure 4 presents the out-of-sample AUCs from models estimated each quarter from 2017 Q1 to 2024 Q4 using the three different sets of predictors. The red line corresponds to models that include only bank characteristics—size, profitability, and balance sheet composition—as predictors. The blue dashed line reflects models that use only the selected BitSight cybersecurity signals identified through the feature selection procedure described earlier. The green dashed line shows the combined specification, incorporating both bank characteristics and selected BitSight features. Each model is trained on all historical data prior to the prediction quarter and evaluated on cyber incidents occurring in the subsequent four quarters.

The results demonstrate that both bank characteristics and cybersecurity signals carry independent predictive content. BitSight-only models achieve relatively strong performance with AUCs generally in the 80–85 range. Bank characteristics alone also yield consistent predictive power. Notably, the combined model consistently outperforms the other two across nearly all quarters, with AUCs clustering around 90 and exhibiting less volatility over time. This pattern suggests that cyber incident risk reflects both latent institutional characteristics (e.g., size, business complexity, risk management capacity) and observable security posture as captured by network-level indicators. The performance improvement from combining both sets of variables highlights the value of integrating traditional supervisory data with external cyber risk signals when assessing banks' digital vulnerability.

Aggregating model performance over time further illustrates these patterns. Figure 5 reports the mean out-of-sample AUC across all quarters for five model specifications, including one that uses lagged incidents—defined as the occurrence of a cyber incident at the bank within the past year—as predictors. The lag-incidents-only model attains a much lower average AUC (68.2), indicating that recent incident history alone has limited predictive value and alleviating concerns that our results are driven solely by persistence of incidents at the same institutions or by reporting concentration.[4] BitSight-only and bank-characteristics-only models achieve average AUCs of 82.5 and 86.2, respectively, while combining them yields 89.8. Adding lagged incidents to the combined model pro-

[4] For work on the propensity of firms to report cyber attacks, see Amir, Levi, and Livne (2018).

duces only a marginal increase to 90.2, suggesting that most predictive power derives from bank characteristics and cybersecurity signals rather than incident history. These results confirm that the two information sets are complementary and that the joint specification consistently delivers the most accurate forecasts of future cyber incidents.

Figure 6 provides a validation of the model's predictive power by examining realized incident rates as a function of out-of-sample predicted risk. For each prediction quarter, we aggregate banks into deciles based on their predicted probability of a cyber incident in the next year (as estimated by the random forest model), and then calculate the actual proportion of banks experiencing an incident within each probability bin. The results reveal a monotonic relationship: as predicted risk increases, the observed incident rate rises sharply.

This pattern is especially pronounced for banks with assets greater than $10 billion. In the top decile of predicted probabilities, nearly 90% of large banks experience a cyber incident in the subsequent year. By contrast, incident rates remain very low among institutions in the lowest deciles. The clear separation across predicted risk bins and size groups underscores both the accuracy and practical utility of the model for supervisory surveillance. The ability to identify those banks most likely to suffer cyber events provides a basis for targeted monitoring and risk management interventions.

## 3.3. Robustness

One important consideration in assessing model performance is whether predictive accuracy varies systematically with bank size. Larger banks differ from smaller institutions in several dimensions relevant to cyber risk—including IT infrastructure complexity, regulatory scrutiny, and public visibility—which may influence both the likelihood of incidents and the observability of relevant security signals. To examine this, we use the same random forest model with both bank characteristics and BitSight features separately for three sub-samples: small banks (assets below $1 billion), mid-sized banks ($1–10 billion), and large banks (above $10 billion). This split allows us to evaluate whether the model's predictive content is robust across the size distribution, rather than being driven disproportionately by one segment of the industry.

The results, shown in Figure 7 indicate that predictive performance remains strong across all size categories, with AUCs consistently above 80 for most periods. Large banks exhibit the highest and most stable AUCs suggesting that the model captures their cyber risk well—likely reflecting richer and more reliable external security signals as well as more frequent incident reporting. Mid-sized banks also show robust performance, with AUCs generally above 85. Small banks, while somewhat more volatile, still achieve AUCs well above chance levels, underscoring that the combined predictor set retains substantial forecasting power even when applied to institutions with fewer observable signals and potentially less complete incident reporting. Overall, the consistency of results across size segments reinforces the general applicability of the model as a supervisory tool for monitoring cyber risk throughout the banking sector.

We also test the robustness of our results to alternative prediction horizons, reestimating the same combined random forest model using forward-looking windows of one quarter, two quarters, and one year. Figure 8 shows that predictive performance remains consistently strong across all horizons, with AUCs generally in the 80–90 range. The one-year horizon yields the most stable and slightly higher performance on average, while shorter horizons show greater quarter-to-quarter volatility—particularly the one-quarter model, which experiences sharper fluctuations. Nevertheless, even at the shortest horizon, the model retains substantial predictive power, indicating that the same set of predictors is effective for both near-term and longer-term cyber risk forecasting. These results confirm that the model's forecasting ability is not tied to a specific horizon, enhancing its practical utility for supervisory monitoring across different planning and intervention timeframes.

## 3.4. Comparison with Textual Metrics

An alternative way to measure exposure to cyber risk is the analysis of related public disclosures by firms, on the basis that firms reveal some information about their exposure to such risk or their ability to prevent incidents. Two prominent examples of this approach are described in Florackis et al. (2023a) (FLMW), who analyze risk-factor information from 10-Ks, and Jamilov et al. (2023) (JRT), who analyze quarterly earnings calls. In this section we compare these metrics to our baseline model combining bank characteristics and BitSight risk-indicators, evaluating each approach against Zywave incidents on a uniform sample of banks. We confirm that both textual metrics have some ability to rank-order banks by incident risk, but our model performs much better and mostly subsumes information in the textual metrics.

Because our sample of banks is predominantly private whereas FLMW and JRT cover public firms spanning many industries, the number of firms in the merged sample is quite small. Scores from JRT match our quarterly frequency, whereas for FLMW we repeat annual scores for each quarter. In addition, overlap in time is limited: our merged sample begins in 2016 and ends in 2018.[5] During this interval the set of banks available from all sources is quite stable, with between 98 and 107 banks depending on the quarter.

We compare FLMW and JRT to two versions of our model. Both follow the fitting and out-of-sample evaluation procedures described previously in this section, using bank characteristics in Table 1 the set of BitSight variables included in Table 2, but applied to the merged sample including FLMW and JRT scores. The second model variant adds the scores of FLMW and JRT alongside other independent variables, to assess incremental performance gains.

Figure 9 shows the results, evaluating performance by quarterly AUC. Although FLMW and JRT generally achieve AUC above 50%, performance is variable and they

[5] BitSight coverage is limited during 2015, the first year for which it is available. We obtain cyber scores for FLMW via the link provided in Florackis, Louca, Michaely, and Weber (2023b), and cyber scores for JRT from Rustam Jamilov at `https://users.ox.ac.uk/~econ0628/Cyber_Risk_Data.zip`.

each fall below 50% in one quarter. FLMW achieves a maximum AUC of around 65% in 2017 Q1 whereas JRT achieves a maximum AUC of around 75% in 2017 Q3. In contrast our baseline model performs better and more consistently, with minimum AUC still above 85% and maximum AUC around 95%. Adding FLMW and JRT scores to our model improves performance slightly, by around 1% on average.

The superior performance of our model may be unsurprising given that it is specialized to the banking sector whereas FLMW and JRT are cross-industry metrics. However, the comparison establishes that these textual metrics contain little information beyond what is already captured by the combination of firm characteristics and BitSight risk-indicators, at least for public banks.

# 4. Analysis of Explanatory Variables and Interpretation of Results

Having established the practicality of our model for out of sample cyber incident prediction, we now investigate the main drivers of the model in more detail. Which predictive variables are the most important in the banking sector? How do they relate to incidents and to each other? Since the focus of this section is interpretation rather than out of sample prediction, we investigate random forest models fitted to our full training sample, from Q1 2015 through Q4 2023 – the same data available in the final quarter analyzed in Section 3.

We begin by characterizing variable importance with all candidate variables included in the model, then consider whether relative importance changes in models where subsets of the variables are excluded, following the flow of Figure 10. Figure 10a compares all independent variables according to two commonly used importance metrics: the mean decrease in Gini impurity and mean decrease in classification accuracy due to permutation. Gini impurity is used internally by the random forest algorithm, to select the variable within a candidate set that maximally decreases heterogeneity of class among observations along each branch of the tree, i.e., splitting banks into those with incidents and those without in our case. Permutation importance captures the reduction in classification accuracy when the values of a given variable are randomly shuffled. For each metric, we normalize the sum of importance across variables to 100. In our case relative variable importance is similar whether impurity or permutation is used, with a couple of notable exceptions.

One result from Figure 10a is that bank characteristics are important to cyber risk. Total assets, deposits/assets, and loans/assets are the three most important variables for predicting incidents, a finding similar to that of Jiang et al. (2024) for public firm characteristics that we extend to private banks. Total assets, in particular, has importance roughly twice that of the most important BitSight variable, which is web application security. ROE and asset growth, the two remaining characteristics, are also in the top 10 candidate variables by impurity but are among the least important variables by permutation, constituting the two cases for which impurity and permutation differ the most. After web application security, the most important BitSight variables are TLS/SSL configuration, DKIM records, TLS/SSL certificates, and patching cadence, consistent with results of our feature selection process towards the end of the sample period. However, BitSight variables potentially exploited, botnet infections, and spam propagation fall into the bottom five candidate variables by importance based on the full training sample, despite being among the most important variables at the beginning of our evaluation period in 2017 (see Table 2).

In Section 4, we evaluated models with bank characteristics and BitSight risk-indicators either separately or in combination. One possibility is that the relative importance of some variables changes once correlated alternatives are included, e.g., that BitSight variables proxy for bank characteristics or vice versa. Figure 10b and Figure 10b show this is not the case, as the relative importance of characteristics and BitSight risk-indicators remains the same for each model variant.[6]

To investigate variable interactions in more detail, Figure 11 shows network visualizations of variable interactions. Node size and color tint indicate Gini impurity, whereas edge linewidth and color tint indicate interaction importance according to the unnormalized H-statistic of Friedman and Popescu (2008). The H-statistic is an unsigned measure of variable interaction based on the mean squared difference between the model prediction when two variables are allowed to interact (joint prediction) or not (synthesis of uni-variate predictions). To reduce figure complexity and consistent with previously discussed results, we focus on interactions within bank characteristics (Figure 11a) and within BitSight risk-indicators (Figure 11b). Throughout our remaining analysis, we also eliminate BitSight risk-indicators never included in our predictive model, per Table 2.

Figure 11a shows that the most important characteristics, such as total assets and loans/assets, also have the strongest interactions with other characteristics on average. However the strongest interaction is between loans/assets and deposits/assets. Although these characteristics have a natural relationship, e.g., as deposits offer a stable funding source for making loans, the interaction as regards cyber incident risk is novel. In combination, the two variables might proxy for a large number of customer relationships, both depositors and lenders, which could make the bank a more attractive target or correspond to an increased attack surface. By contrast, the smallest interactions are between asset growth and ROE and asset growth and deposits/assets.

Figure 11b shows interactions between selected BitSight variables. Since total assets is overall the most important variable, we also plot interactions between it and the BitSight variables; such interactions are, however, very small. In contrast to the results in Figure 11a for characteristics, the most important BitSight interactions do not generally originate from the most important variable, but rather from patching cadence, which rounds out the top five BitSight variables by importance. The only large interaction not involving

[6] Of course, the absolute importance of each variable does drop when the variable set is expanded: the figures normalize importances within each subset to 100, to emphasize relative values.

patching cadence is between DKIM records and TLS/SSL certificates. Patching cadence interacts strongly with all four other BitSight variables selected during the final part of the evaluation period: web app. security, TLS/SSL configuration, TLS/SSL certificates, and DKIM records. A natural interpretation is that the risk of an incident increases when there is both unpatched software and a weakness in communication protocols.

We now investigate in detail the relationship between each of our selected variables and incident rates as uncovered by the random forest model, via partial dependency plots (PDP). Since random forests allow for general non-parametric dependencies, the relationship between each predictive variable and incidents is most easily summarized by plotting the model-implied incident probability conditional on the value of a selected variable, taking expectations over other variable values according to the sample distribution.

Figure 12 shows the PDP for each bank characteristic, with annual incident probabilities on a log scale on the y axes. In the top left plot, we see incident rates rising sharply with the log of total assets, particularly in the transition between mid-sized and large banks, consistent with the higher incident rate among large banks illustrated in Figure 3. However, the PDP plot reveals a relationship between total assets and incidents that is relatively flat within large banks and within small banks. The remaining PDP plots for bank characteristics are, by contrast, unambiguously non-monotonic. Deposits/assets and loans/assets are similar ”smirks” varying over a wide range, falling from approximately 15% annual incident probability for very small x-values to approximately 3% for middle x-values before rising again to around 7% for large x-values. Asset growth and ROE share similar ”smile” patterns, but vary over a narrow range of around 3-5% annual incident probability. In short, banks that are near the extremes in terms of their growth rate, asset composition, or liability composition are at higher risk of cyber incidents.

Figure 13 shows the PDP for each selected BitSight risk-indicator. First, we note that variation in incident probabilities is less for BitSight variables than we observed for bank characteristics, with the exceptions of ROE and asset growth. However, most BitSight variables have the expected relationship to incident probabilities – rising as the scores fall – with incident probabilities increasing by at least 50% and in some cases doubling conditional on very low scores. Very low scores are rare, but they are observed, and incidents are also uncommon for small-and-medium-sized banks.[7]

However, Figure 13 includes a few curiosities worthy of remark, as not all relationships are monotonic or of the expected sign. Web application security and TLS/SSL certificates both show incident probabilities counterintuitively increasing with scores in the upper range, albeit modestly. This could reflect maximum scores (typically 820) as default values when no information about the risk-indicator is available, such that incident probabilities match the higher unconditional mean for very high scores versus a lower conditional mean with measurably good-but-not-perfect scores. Regardless of the underlying mechanism, the random forest shines in dealing with such local non-monotonicity relative to more traditional linear models. Although we do not include results here, alternatives such as logistic regression generally find weaker predictive relationships, sometimes with point estimates of opposite sign, precisely because so many observations fall near the top end of the score range, where variation is either uninformative or, worse, blurs the distinction between confirmation of strong security and ”no data.” Since methods such as logistic regression must assign a single coefficient summarizing the global relationship, murky relationships in small but frequently observed score ranges may dominate.

Finally, there is one variable in Figure 13 where high scores are unambiguously and counterintuitively ”bad news”: DKIM records, in the top-right plot. Although we lack a conclusive explanation, we do find that the DKIM records score increases following an incident, perhaps because it is an easy risk-indicator for firms to improve in response.

[7] That being said, botnet infections, potentially exploited, and spam propagation have relatively low cross-sectional or time variation in general, which is one reason why these variables are not selected features for much of our evaluation period.

Since lagged-incidents are predictive of future incidents, this could explain why high DKIM scores are associated with increased probability of an incident.

Before concluding, we revisit in more detail the most interesting set of interactions between BitSight variables. Figure 14 shows pair-wise interaction heatmaps for the five most important BitSight variables on the upper triangle, reprises univariate PDP plots on the diagonal, and has scatterplots of observations colorized by model-implied incident probability in the lower triangle. The figure serves two main purposes. First, recall that patching cadence had the strongest interactions with other variables in Figure 11b, but the H-statistic is unsigned, leaving the nature of the interaction ambiguous. The heatmaps show that the model predicts the highest incident probabilities when patching cadence is low *and* TLS/SSL certificates, TLS/SSL configuration, or web application headers is low. This is consistent with our earlier interpretation that unpatched software combines with weak communication protocols to increase incident risk. Second, the scatterplots show that, although moderate and low scores are less common than high scores, they are observed not only for individual variables but also for several combinations of variables. Such combinations may be critical to identifying high incident risk.

## 5. Conclusion

This paper develops a sector-wide, forward-looking measure of cyber risk exposure for the U.S. banking industry, combining externally observed cybersecurity posture, realized incident data, and bank characteristics. By integrating BitSight's granular technical risk indicators with regulatory Call Reports and Zywave's incident records, we construct a bank-quarter panel covering institutions of all sizes from 2015 to 2024. Using a dynamic feature-selection framework within a random forest model, we show that both cybersecurity posture and financial characteristics provide independent and complementary predictive power for cyber incidents over the subsequent year.

Our results are robust across bank sizes and prediction horizons, underscoring the model's applicability for supervisory monitoring throughout the sector. The analysis reveals that predictive accuracy does not rely on persistence in which banks experience incidents, but instead reflects contemporaneous differences in network security practices, IT configuration hygiene, and institutional fundamentals. Moreover, by identifying the specific technical risk vectors most closely associated with elevated incident probabilities—such as patching cadence, TLS/SSL configuration, and web application security—the framework yields actionable insights for targeted oversight and internal risk management.

From a policy perspective, the findings highlight the value of integrating commercial cyber risk indicators with supervisory data to generate timely, institution-specific measures of digital vulnerability. Such measures can support microprudential interventions at the bank level as well as macroprudential assessments of systemic exposure to cyber threats.

# References

Ahnert, T., M. Brolley, D. A. Cimon, and R. Riordan (2024). Cyber risk and security investment. *Available at SSRN 4057505*.

Amir, E., S. Levi, and T. Livne (2018). Do firms underreport information on cyber-attacks? evidence from capital markets. *Review of Accounting Studies 23*(3), 1177–1206.

Berger, A. N., F. Curti, A. Mihov, and J. Sedunov (2022). Operational risk is more systemic than you think: Evidence from us bank holding companies. *Journal of Banking & Finance 143*, 106619.

Bouveret, A. (2018). Cyber risk for the financial sector: A framework for quantitative assessment. (18/143).

Caro Rincon, A. and G. Ordóñez (2023). The impact of cyber security management practices on the likelihood of cyber events and its effect on financial risk. Technical report, Moody's Analytics.

Chernobai, A., A. Ozdagli, and J. Wang (2021). Business complexity and risk management: Evidence from operational risk events in us bank holding companies. *Journal of Monetary Economics 117*, 418–440.

Choi, S. J. and M. E. Johnson (2021). The relationship between cybersecurity ratings and the risk of hospital data breaches. *Journal of the American Medical Informatics Association 28*(10), 2085–2092.

Duffie, D. and J. Younger (2019). *Cyber runs*. Brookings.

Eisenbach, T. M., A. Kovner, and M. Lee (2025). When it rains, it pours: Cyber vulnerability and financial conditions. *Economic Policy Review 31*(1), 1–24.

Eisenbach, T. M., A. Kovner, and M. J. Lee (2022). Cyber risk and the us financial system: A pre-mortem analysis. *Journal of Financial Economics 145*(3), 802–826.

Eling, M. and J. Wirfs (2019). What are the actual costs of cyber risk events? *European Journal of Operational Research 272*(3), 1109–1119.

Florackis, C., C. Louca, R. Michaely, and M. Weber (2023a). Cybersecurity risk. *The Review of Financial Studies 36*(1), 351–407.

Floracki, C., C. Louca, R. Michaely, and M. Weber (2023b). Cybersecurity risk: The data. *Chicago Booth Research Paper* (23-01).

Friedman, J. H. and B. E. Popescu (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics 2*(3), 916–954.

Gogolin, F., I. Lim, and F. Vallascas (2021). Cyberattacks on small banks and the impact on local banking markets. *Available at SSRN 3823296*.

Heo, Y. (2024). Cybersecurity and bank fragility. *Available at SSRN 4660090*.

Huang, H. H. and C. Wang (2021). Do banks price firms' data breaches? *The Accounting Review 96*(3), 261–286.

Jamilov, R., H. Rey, and A. Tahoun (2023). The anatomy of cyber risk. Technical report, National Bureau of Economic Research.

Jiang, H., N. Khanna, Q. Yang, and J. Zhou (2024). The cyber risk premium. *Management Science 70*(12), 8791–8817.

Jin, Y. J., N. Li, S. Liu, and S. M. K. Nainar (2023). Cyber attacks, discretionary loan loss provisions, and banks' earnings management. *Finance Research Letters 54*, 103705.

Kamiya, S., J.-K. Kang, J. Kim, A. Milidonis, and R. M. Stulz (2021). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics 139*(3), 719–749.

Kashyap, A. K. and A. Wetherilt (2019). Some principles for regulating cyber risk. In *AEA Papers and Proceedings*, Volume 109, pp. 482–487. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Kopp, E., L. Kaffenberger, and C. Wilson (2017). Cyber risk, market failures, and financial stability. IMF Working Paper 17/185, International Monetary Fund.

Murphy, A., M. L. Tindall, K. Klemme, J. I. Suek, and S. J. Dunbar (2025, May). What drives cyber losses at u.s. banks? potential statistical markers. Working Paper 2520, Federal Reserve Bank of Dallas. Accessed: 2025-08-15.

Ottonello, G. and A. E. Rizzo (2024). Do software companies spread cyber risk? *Available at SSRN*.

Ramírez, C. A. (2025). On equilibrium cyber risk. *Economics Letters 251*, 112307.

Sheneman, A. (2025). Cybersecurity risk and bank loan contracting. *Available at SSRN 3406217*.

The International Monetary Fund (2024, April). Cyber risk: A growing concern for macrofinancial stability. In *Global Financial Stability Report*, Chapter 3. International Monetary Fund. Accessed: 2025-08-15.

### Figure 1: BitSight Signal Distributions

This figure presents kernel density plots of BitSight cybersecurity indicator scores for U.S. banks, grouped by total asset size: less than $1 billion (green), $1–$10 billion (blue), and greater than $10 billion (orange). Each panel corresponds to one of 15 indicators, including measures of malicious activity, patching practices, protocol configuration, and software security. The x-axis shows the BitSight score, and the y-axis shows the estimated density within each size category.

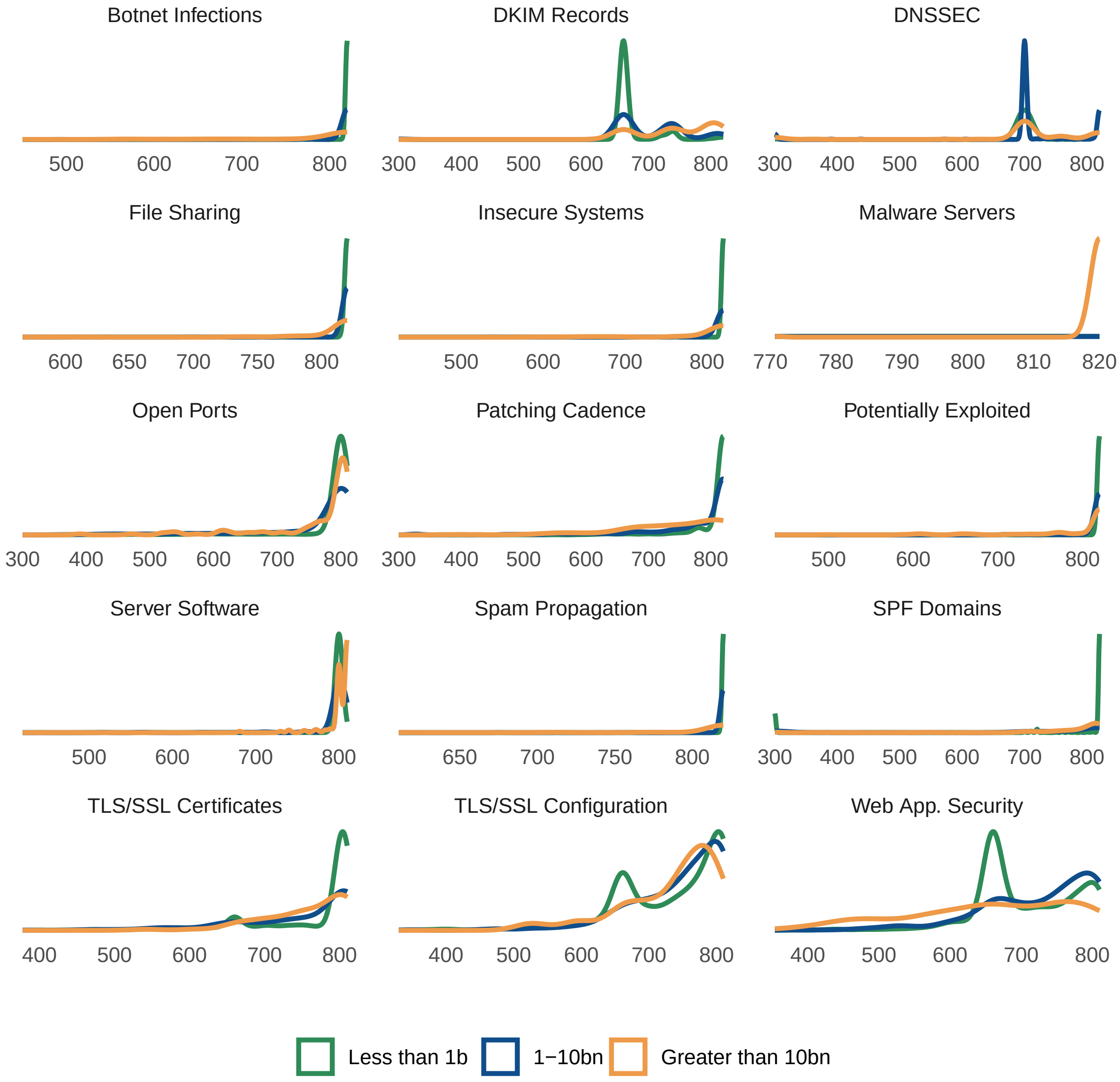


**Figure 2: Number of incidents at banks**

Quarterly counts of reported cyber incidents at U.S. banks from 2015 to 2024, disaggregated by incident category: data–malicious breach, identity–fraudulent use/account access, IT–configuration/implementation errors, IT–processing errors, network/website disruption, and phishing/spoofing/social engineering.

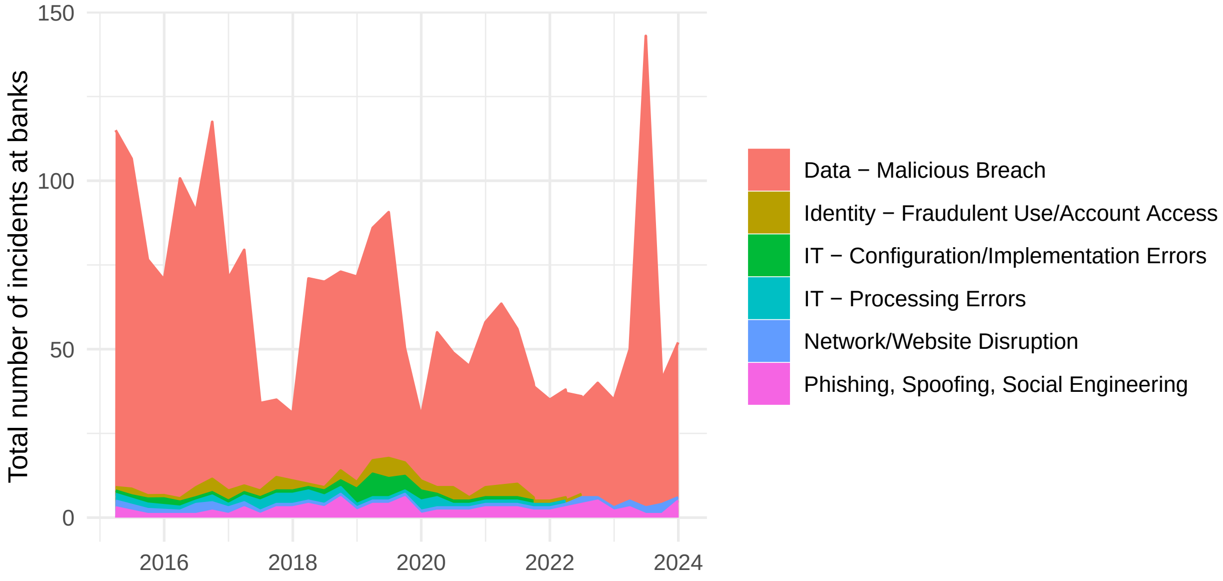
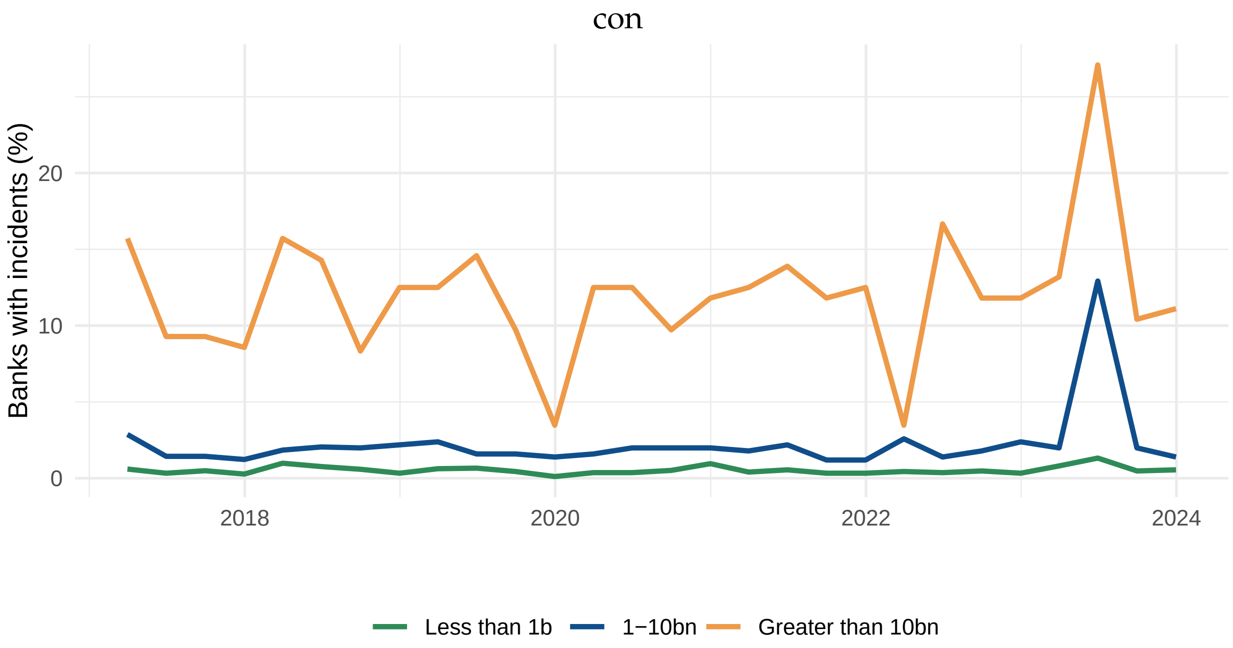


**Figure 3: Percent of banks with incidents**

Time series of the percentage of banks experiencing at least one cyber incident, segmented by asset size categories (less than $1 billion, $1–10 billion, and greater than $10 billion), from 2017 to 2024.



## Table 1: Bank Characteristics

This table reports summary statistics for banks in and out of the estimation sample, grouped by asset size: greater than 10 billion, between 1 and 10 billion, and less than 1 billion. For each group, the table shows the number of banks ($N$), mean asset growth, deposits-to-assets ratio, loans-to-assets ratio, return on equity (ROE), and total assets (in thousands of USD).

| | Greater than 10bn | | 1 - 10bn | | Less than 1bn | |
|---|---|---|---|---|---|---|
| | Not in Sample | In Sample | Not in Sample | In Sample | Not in Sample | In Sample |
| N | 39 | 101 | 154 | 505 | 1,692 | 2,736 |
| Asset Growth | 2.029 | 2.761 | 3.774 | 2.796 | 1.826 | 1.973 |
| Deposits/Assets | 0.754 | 0.762 | 0.805 | 0.815 | 0.806 | 0.833 |
| Loans/Assets | 0.664 | 0.659 | 0.687 | 0.720 | 0.621 | 0.653 |
| ROE | 10.853 | 10.659 | 9.457 | 10.958 | 6.584 | 8.978 |
| Total Assets | 75,313,255 | 117,259,077 | 2,465,458 | 2,732,737 | 224,905 | 284,449 |

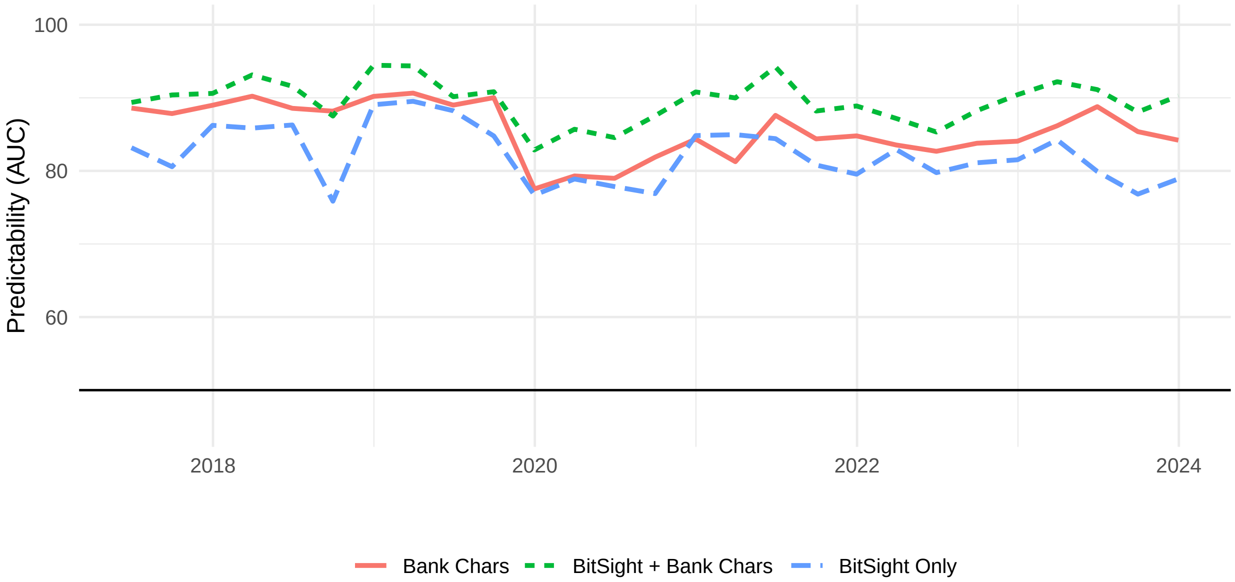
## Table 2: Selected Features

This table lists the BitSight cyber risk signals selected by the model in each period between 2017Q2 and 2023Q4. For each date range, the table shows the set of features retained after the model's feature selection process, which was applied quarterly. The listed indicators include measures of system vulnerabilities, security configurations, and potential exploit exposures, with some features appearing consistently across multiple periods and others appearing only in specific intervals.

| From | To | Selected Features |
|---|---|---|
| 2017 Q2 | 2017 Q4 | Botnet Infections, DKIM, Potentially Exploited, Spam Propogation, TLS/SSL Configuration |
| 2018 Q1 | 2018 Q2 | Botnet Infections, DKIM, Potentially Exploited, TLS/SSL Configuration, Web Application Security |
| 2018 Q3 | 2022 Q3 | DKIM, Patching Cadence, Potentially Exploited, TLS/SSL Configuration, Web Application Security |
| 2022 Q4 | 2023 Q4 | DKIM, Patching Cadence, TLS/SSL Certificates, TLS/SSL Configuration, Web Application Security |

## Figure 4: Out-of-Sample Predictability - Quarterly

The figure plots quarterly out-of-sample predictability, measured by AUC, for three model specifications: Bank Characteristics (solid red line), BitSight + Bank Characteristics (dashed green line), and BitSight Only (dashed blue line), over the period from 2017 to 2024.



## Figure 5: Out-of-Sample Predictability - Mean

This figure presents mean out-of-sample predictability (AUC) for five model specifications. Models using only lagged incidents achieve the lowest predictive accuracy, while those combining BitSight signals and bank characteristics perform substantially better. The inclusion of lagged incidents alongside BitSight and bank characteristics produces the highest mean AUC, marginally improving over the combined BitSight–bank characteristics model. These results provide a clear comparison of the relative contribution of different predictor sets to model performance.

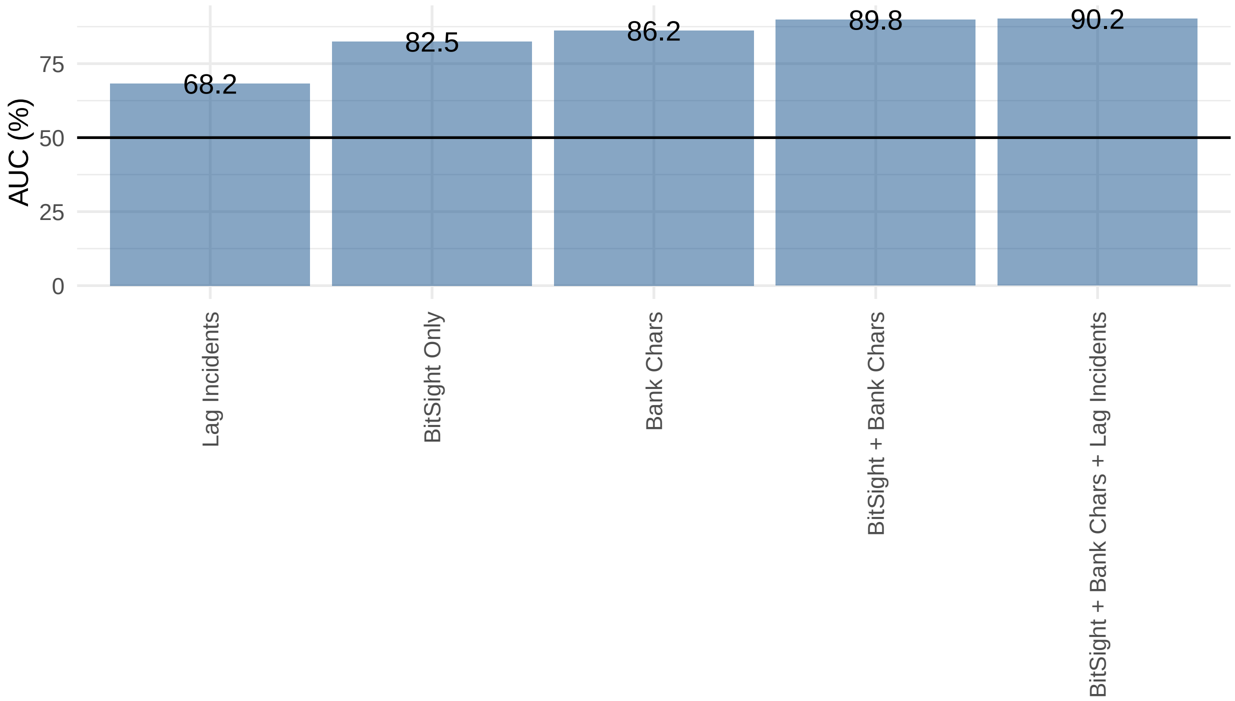


**Figure 6: Out-of-Sample Predictability - Mean**

This figure presents the distribution of actual cyber incidents across predicted probability deciles, segmented by bank size. The x-axis groups banks into deciles based on their out-of-sample predicted probability of a cyber incident, and the y-axis reports the observed percentage of banks within each decile that experienced an incident. Separate series are shown for banks with assets less than 1 billion, between 1 and 10 billion, and greater than 10 billion.

The plot allows for comparison of model calibration and discrimination across size categories, highlighting differences in incident prevalence between low- and high-risk deciles. It also facilitates visual assessment of how predicted risk translates into realized outcomes within each asset class over the probability spectrum.

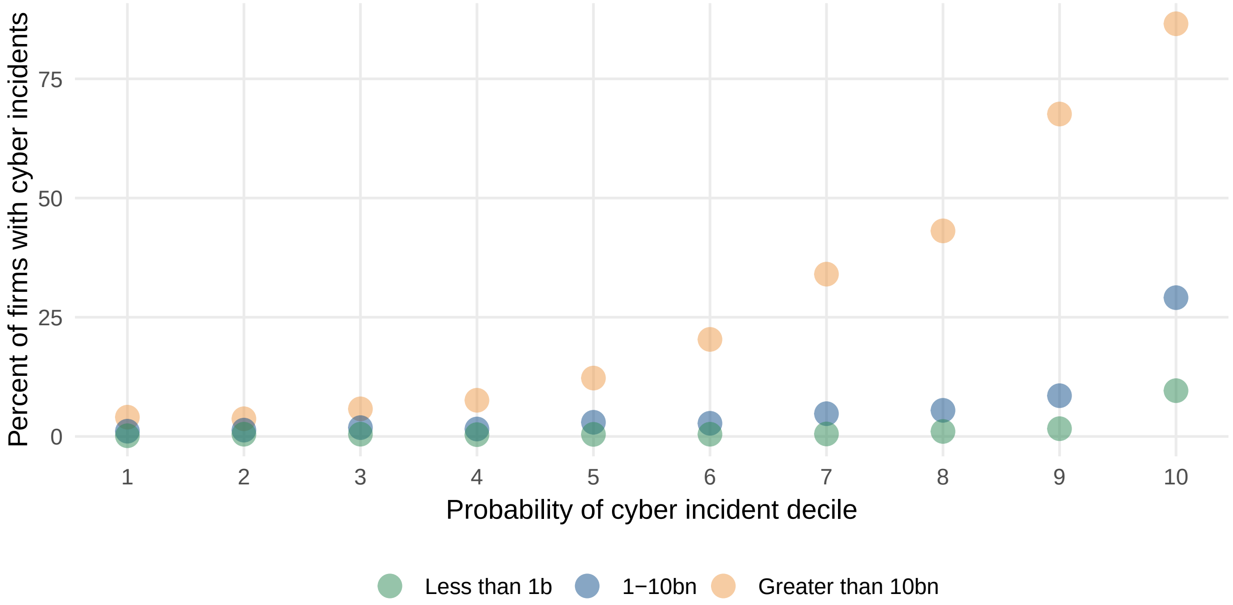


**Figure 7: Out-of-Sample Predictability - By Size**

This figure plots quarterly out-of-sample predictability (AUC) from 2017 to 2024 separately for banks with total assets less than $1 billion, $1–10 billion, and greater than $10 billion. Each line represents a size category: green for less than $1 billion, blue for $1–10 billion, and orange for greater than $10 billion.

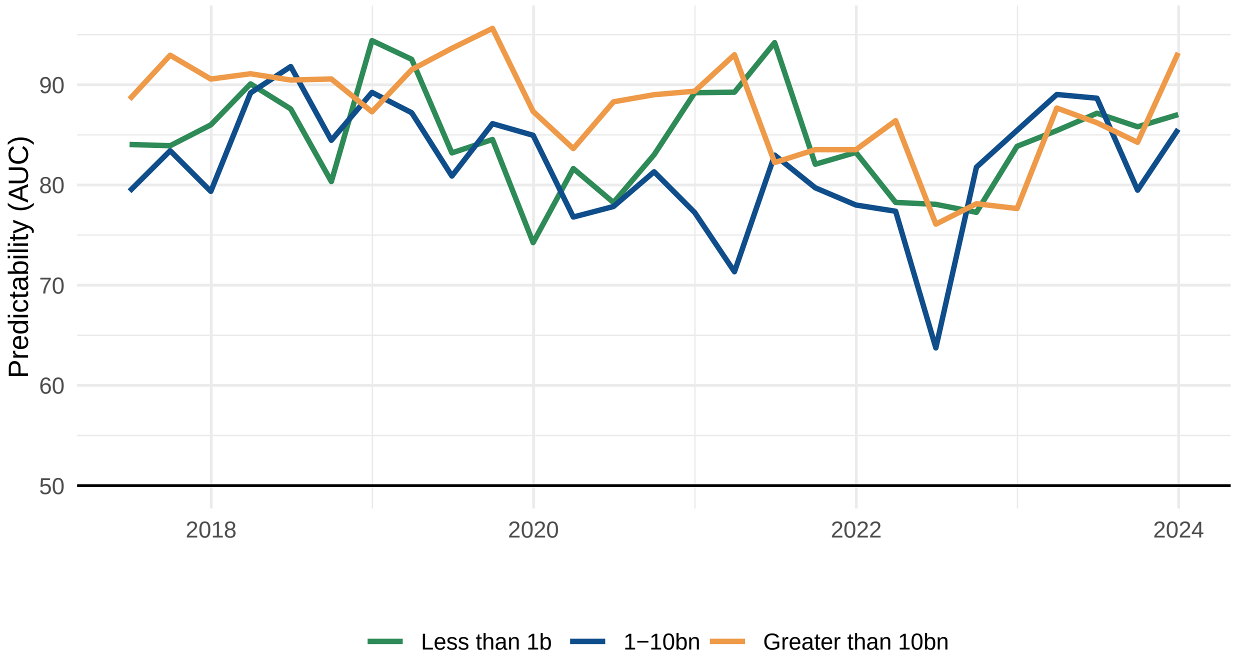


**Figure 8: Out-of-Sample Predictability - By Incident Horizon**

This figure presents the out-of-sample predictability (AUC) from 2017 to 2024, broken down by prediction horizon. The solid red line corresponds to a one-quarter horizon, the dashed blue line to two quarters, and the dashed green line to one year. The horizontal axis marks calendar years, while the vertical axis measures predictability in percentage AUC terms.

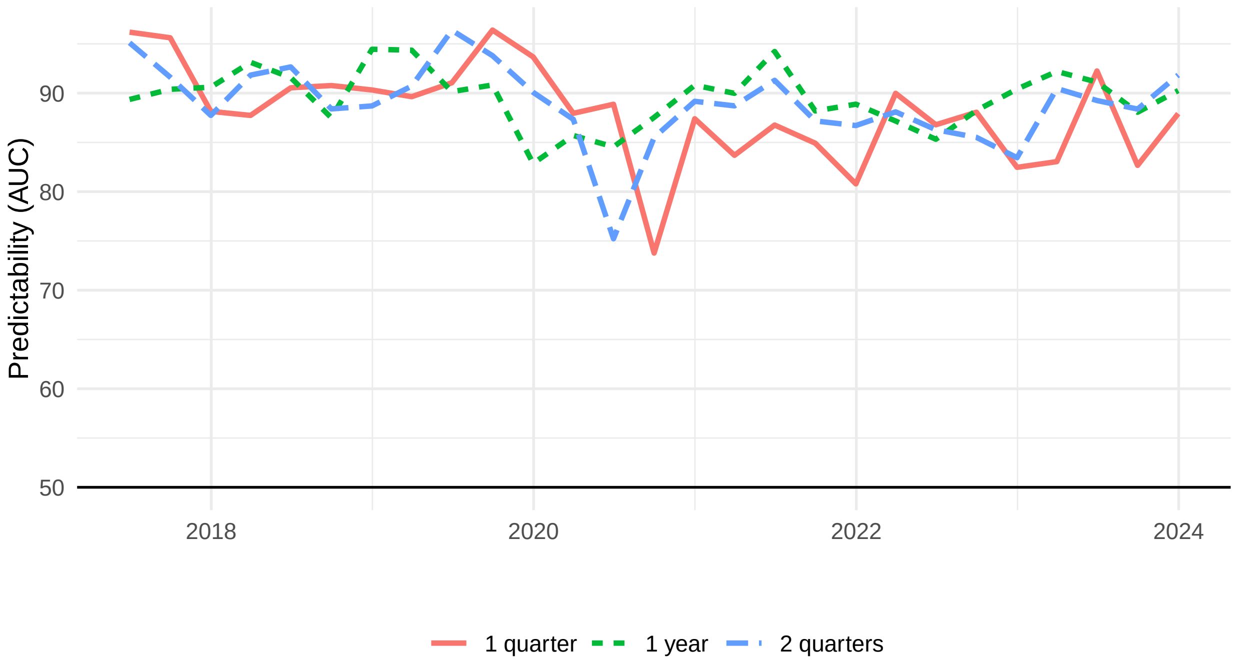
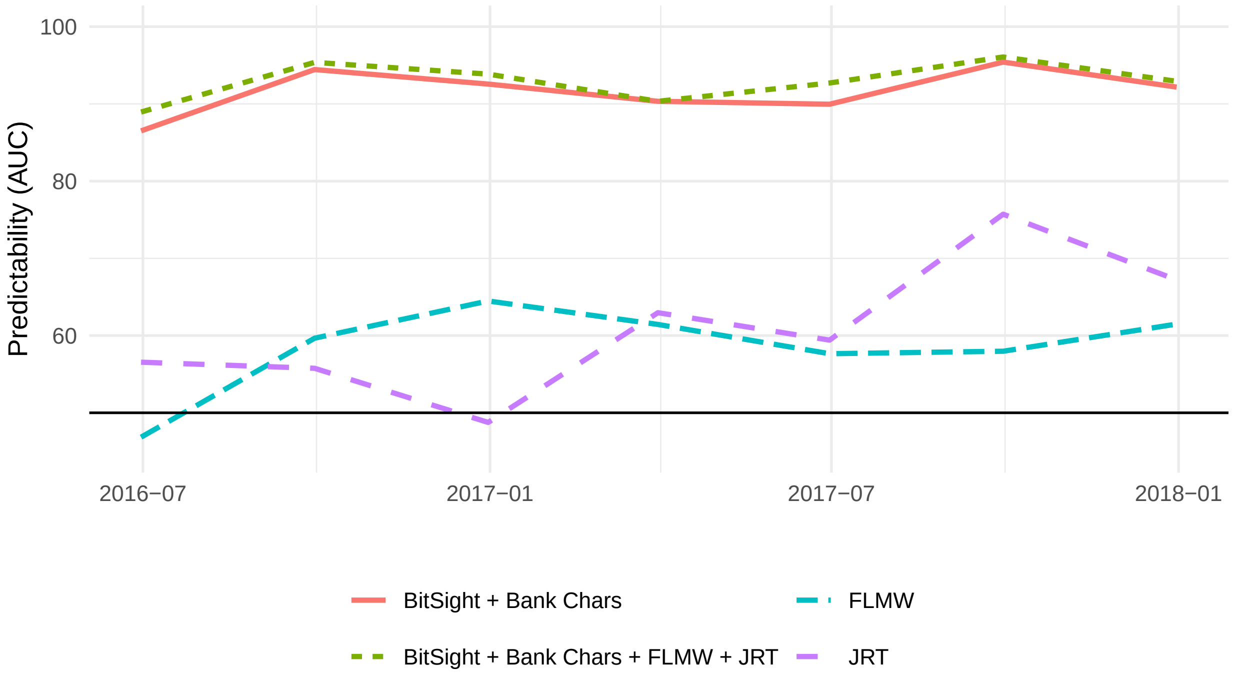


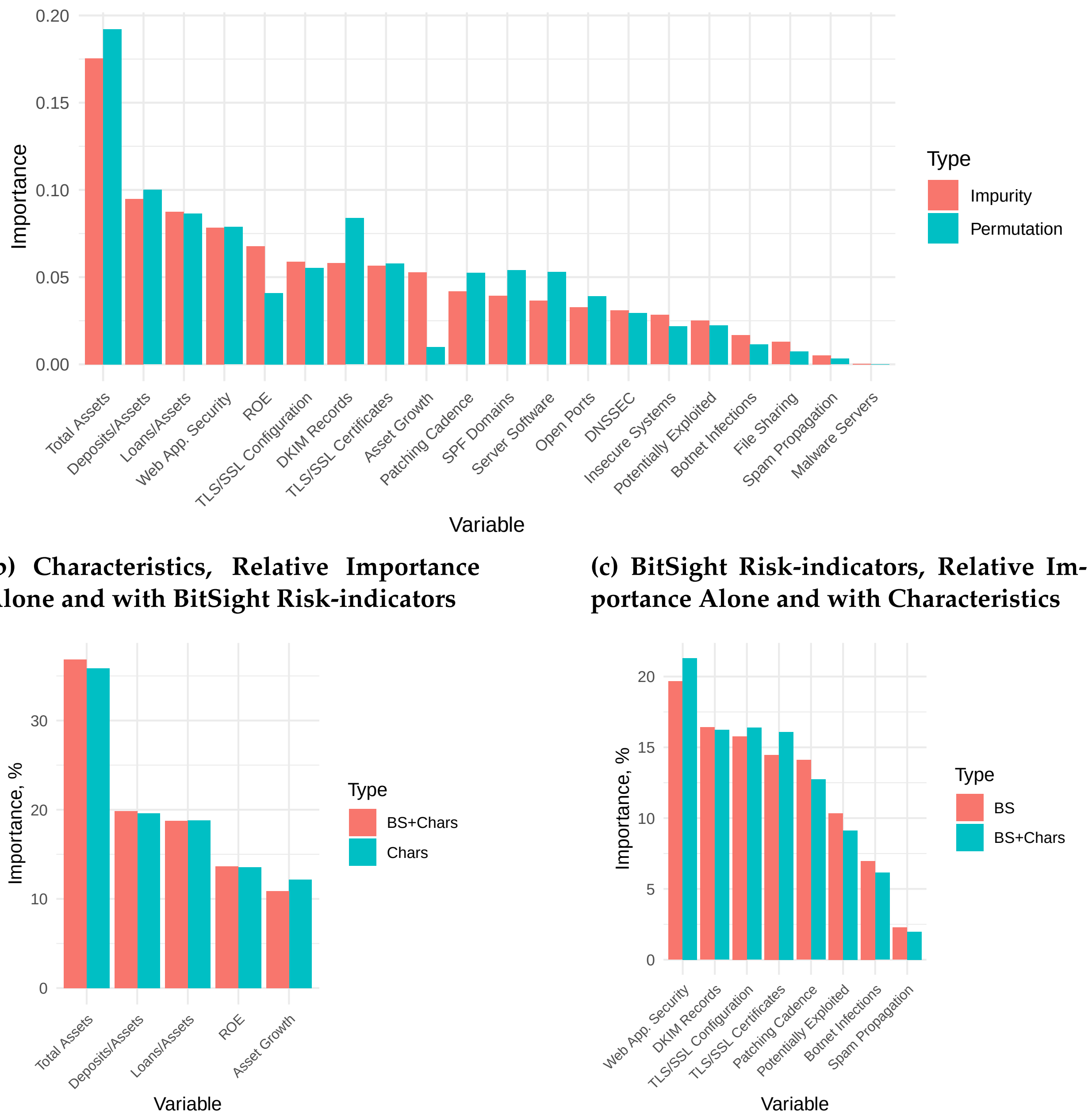
**Figure 9: Out-of-Sample Predictability - Comparison with Text Metrics**

The figure compares our baseline model, BitSight + Bank Characteristics, with two alternative metrics based on textual analysis, developed in Florackis et al. (2023a) (FLMW) and Jamilov et al. (2023) (JRT) respectively. Performance is assessed using AUC out-of-sample, based on incident occurrence over the following year. Also shown are results for the baseline model combined with textual metrics. Although FLMW and JRT achieve AUC of around 60% on average, indicating better than random performance, our baseline model performs much better, with average AUC above 90%. Adding textual metrics to our model slightly improves performance, by about 1% on average.



## Figure 10: Importance of Independent Variables

This figure presents the variable importance results from the predictive model in three panels. Panel (a) shows the relative ranking of all predictors using both impurity-based and permutation measures. Panel (b) compares the contribution of firm characteristics alone with their contribution when combined with BitSight risk indicators, while Panel (c) compares the contribution of BitSight risk indicators alone with their contribution when combined with firm characteristics. Color coding distinguishes the different variable sets and importance measures across panels, providing a visual comparison of their relative influence in the model.



**(a) All Variables**

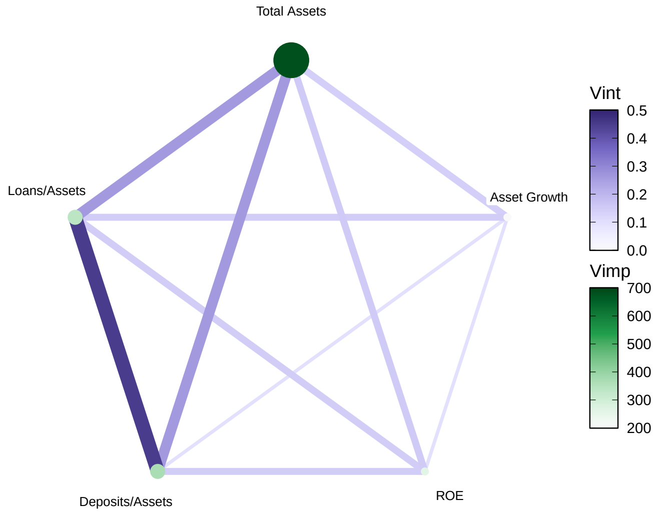
**(b) Characteristics, Relative Importance Alone and with BitSight Risk-indicators**

**(c) BitSight Risk-indicators, Relative Importance Alone and with Characteristics**

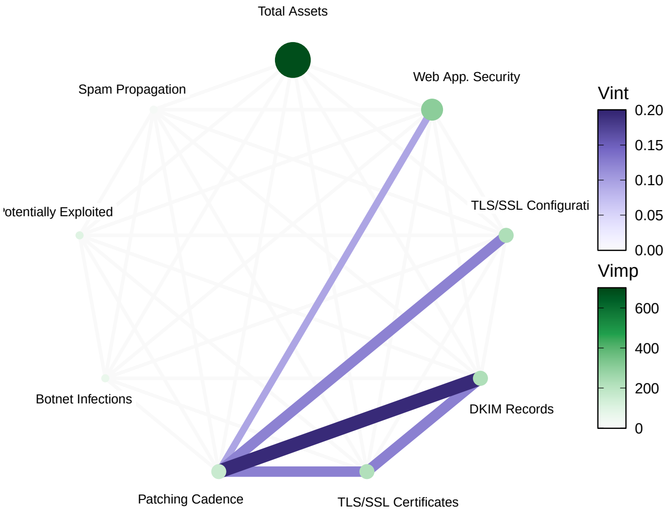
## Figure 11: Variable Importance and Interaction Network

This figure presents network diagrams illustrating the importance and interactions of variables within two groups: bank characteristics (panel a) and BitSight risk indicators (panel b). Each diagram represents variables as nodes, with node size reflecting variable importance and edge thickness representing the strength of interaction between variables. The color gradients correspond to the magnitude of variable importance (Vimp) and interaction strength (Vint). Panel (a) displays the relationships among bank characteristics, while panel (b) shows the relationships among risk indicators, with both highlighting the most influential variables and strongest interactions in their respective groups.

### (a) Bank Characteristics

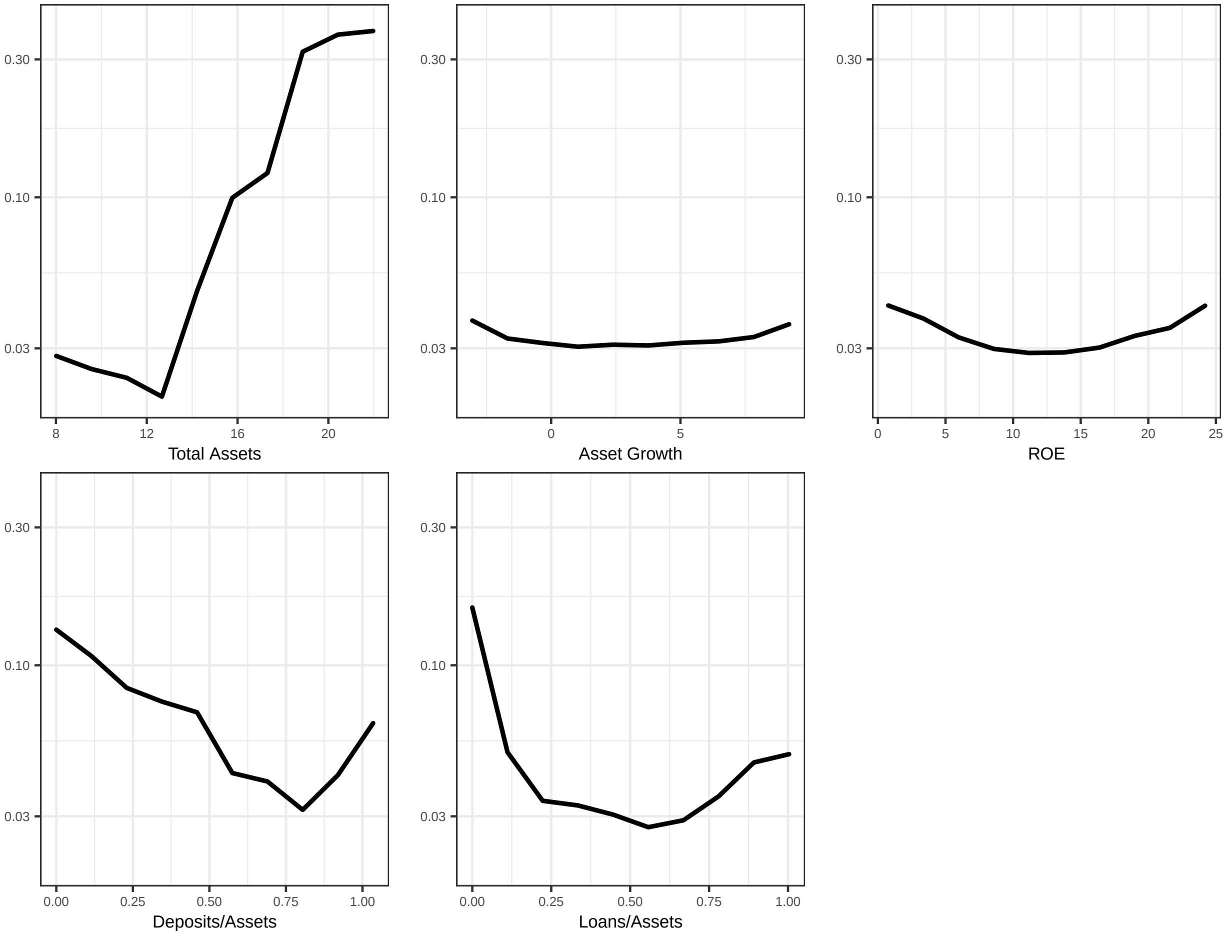


### (b) BitSight Risk-indicators



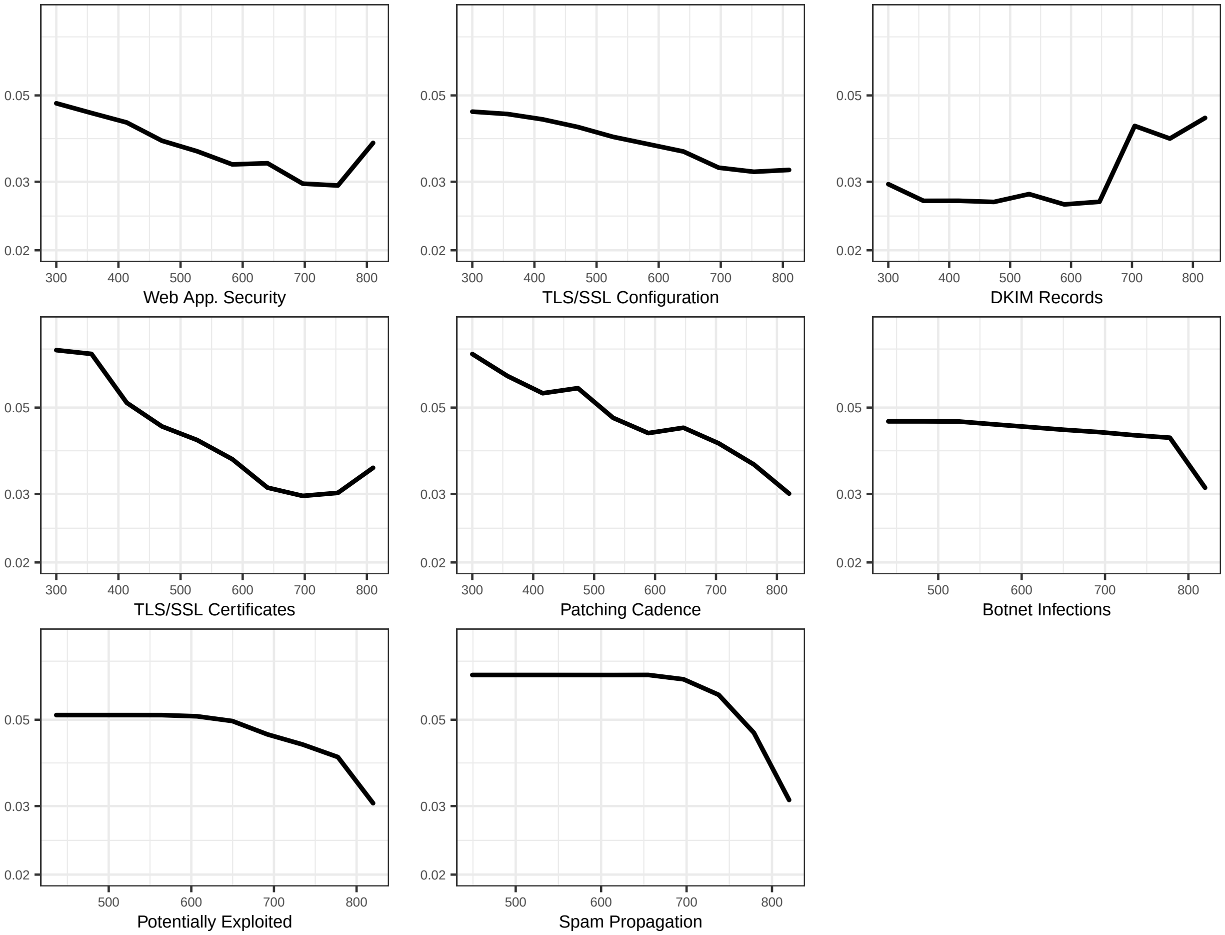
### Figure 12: Partial Dependence Plots – Bank Characteristics

This figure presents partial dependence plots for the five bank characteristics used in the analysis. Each panel isolates the marginal effect of a single characteristic on the predicted outcome, holding other variables constant. The horizontal axis in each plot represents the range of observed values for the characteristic, and the vertical axis shows the corresponding change in the predicted probability scale. These plots illustrate the functional form of the relationship between each bank characteristic and the model's prediction.



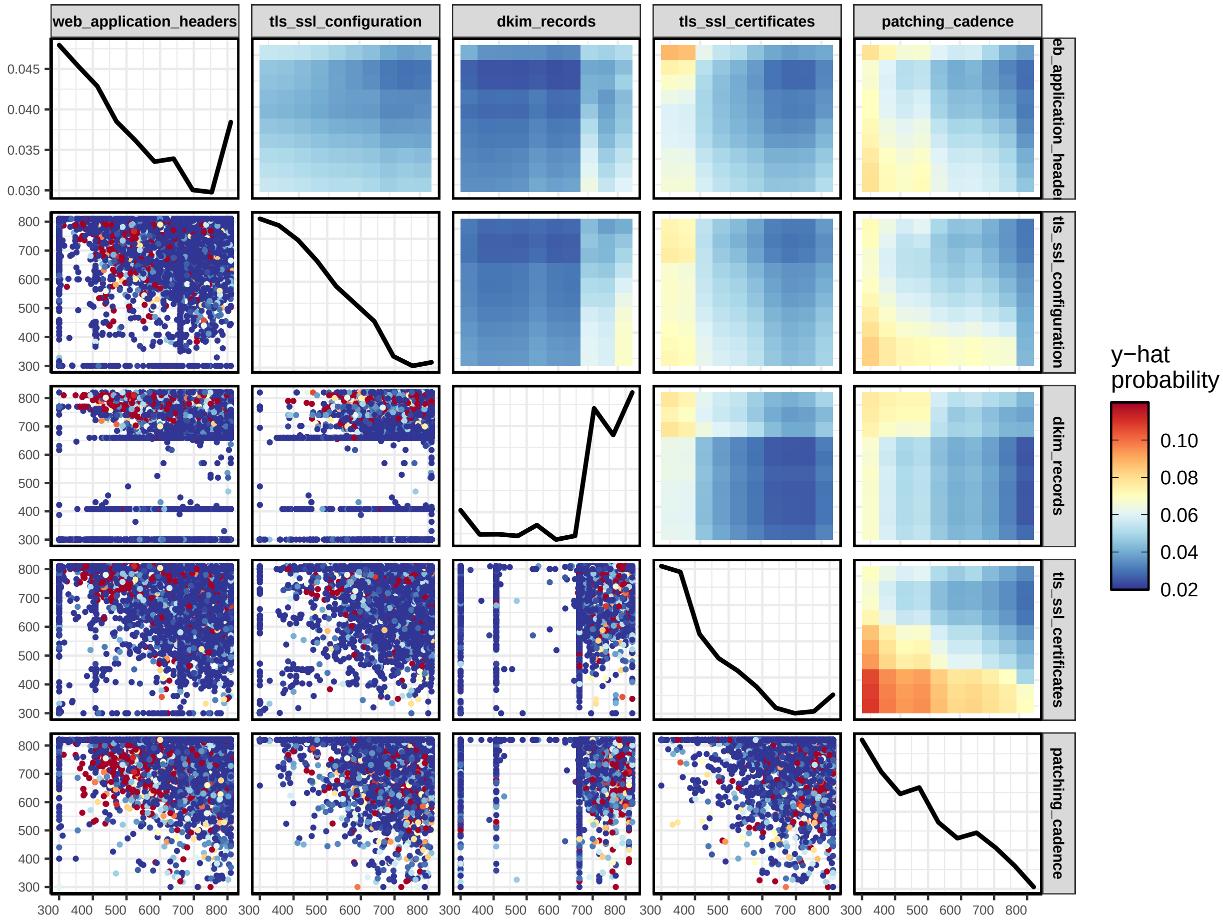
### Figure 13: Partial Dependence Plots – BitSight Risk-indicators

This figure presents partial dependence plots for the BitSight risk-indicators. Each panel shows the relationship between a given risk-indicator and predicted incident likelihood, holding other variables constant. The horizontal axis displays the range of the respective indicator values, while the vertical axis shows the associated partial dependence.



### Figure 14: Pairwise Dependencies – Selected BitSight Risk-indicators

This figure displays pairwise partial dependence plots for selected BitSight risk-indicators. The diagonal panels show the individual partial dependence of each variable on the predicted probability. The upper triangle presents heatmaps of predicted probabilities for each pair of variables, while the lower triangle contains scatter plots of the underlying data points colored by predicted probability. Color gradients in the heatmaps range from blue (lower predicted probability) to red (higher predicted probability), illustrating how predicted outcomes vary across the joint distribution of the selected indicators.


web_application_headers
tls_ssl_configuration
dkim_records
tls_ssl_certificates
patching_cadence
eb_application_heade
tls_ssl_configuration
dkim_records
tls_ssl_certificates
patching_cadence
y−hat
probability